DOCUMENT RESUME

ED 243 307 FL 014 311

AUTHOR Parish, Charles; Perkins, Kyle
TITLE Using Tests of Anaphoric Reference in ESL Reading.
PUB DATE Mar 84
NOTE 18p.; Paper presented at the Annual Meeting of the Teachers of English to Speakers of Other Languages (18th, Houston, TX, March 6-11, 1984).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Cloze Procedure; Comparative Analysis; Connected Discourse; *English (Second Language); Form Classes (Languages); Language Processing; Language Tests; Multiple Choice Tests; *Pronouns; *Reading Comprehension; *Reading Instruction; *Second Language Learning; Test Reliability; Test Validity

ABSTRACT

This study investigates the extent to which beginning to advanced students of English as a second language (ESL) understand antecedent/anaphora relationships in written discourse. The findings are related to previous research on anaphoric reference in first and second language contexts, emphasizing the significance of this factor in measurement of language acquisition. The instruments used for elicitation were two Tests of Anaphoric Reference-Multiple Choice and two Tests of Anaphoric Reference-Cloze Format. Subjects were 50 full-time college-level ESL students with eight different native languages. Five factors of comprehension of anaphoric reference were studied: (1) the relevance of students' proficiency levels on test performance, (2) recognition versus production performance, (3) the effect of numbers of antecedents, (4) the effect of types of pronouns tested, and (5) the effects of the distance between anaphora and antecedent. Results indicate that for three of the four separate tests used, proficiency level was a significant factor; production was more difficult than recognition; the number of antecedents did not affect pronoun assignment; a significant difference in performance was found on the cloze tests but not on the multiple choice tests for three types of pronoun; and distance between anaphora and antecedent made a difference with cloze tests but not multiple choice tests. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED243307

Using Tests of Anaphoric Reference in ESL Reading

Charles Parish

Southern Illinois University

Kyle Perkins

Southern Illinois University

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Charles Parish
Kyle Perkins
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Abstract

This paper investigates the extent to which beginning to advanced ESL reading students understand antecedent/anaphora relationships in written discourse, and by relating the findings to previous research on anaphoric reference with L1 and L2 contexts, continues the attempt to establish the significance of this factor in the measurement of language acquisition. The elicitation instruments, Tests of Anaphoric Reference--Multiple-Choice (TAR-MC) and Tests of Anaphoric Reference--Cloze Format (TAR-C), were developed by Smith (1973) and described in Fagan, Cooper and Jensen (1975). Normative, validity, and reliability data are reported. Five research questions are posed and answered: the relevance of proficiency levels, recognition vs. production performance, the effect of numbers of antecedents, differential performance with varying pronoun types, and the effect of number of propositions separating pronoun and antecedent.

FL014311

Introduction

Anaphora is a linguistic device of syntactic cross-reference whereby words or groups of words can be used as a surrogate for a preceding word or group of words. Anaphora and anaphoric relations have been studied extensively from different perspectives. Garvey, Caramazza, and Yates (1976) and Caramazza, Grober, and Garvey (1977) studied the role that semantic

factors play in the assignment of pronoun antecedents; they found that the direction of causality assigned to root verbs or implicit causality is a major determinant of pronoun assignment. Fishman (1978) examined the effects of anaphoric references on paragraph comprehension.

Anaphoric processes have been examined from a generative grammar perspective. Hankamer and Sag (1976:392) claimed that there is a difference between syntactically and pragmatically controlled anaphora:

> "Deep" anaphora, which allows pragmatic control and has other properties indicating that the anaphoric relation is determined at an essentially presyntactic level; and another, which we will ultimately call "surface" anaphora, which requires a coherent syntactic antecedent in surface structure and otherwise behaves as a purely superficial syntactic process.

Other investigators have examined the relative difficulty of various types of anaphora. Bormuth, Carr, Manning, and Pearson (1970) studied fourth-grade children's comprehension of 14 different types of anaphora. They found evidence for a hierarchical ranking of pronouns, and the personal pronouns were found to be among the most difficult. In a partial replication of the Bormuth et al. (1970) study, Lesgold (1974) found that Bormuth et al.'s difficulty ordering was not stable. Lesgold reported that in his study the personal pronouns were relatively easy. Richek (1976-77) reported that comprehension of anaphoric forms was affected by the question and sentence frames and sentence complexity that she used in the elicitation materials, but kernels, length, and parallelism did not affect difficulty. A common thread running through the Bormuth et al.

(1970), Lesgold (1974), and Richek (1976-77) studies is that L1 elementary school children exhibit unexpectedly low performance in handling the assignment of pronoun reference and that there is a high magnitude of variability among subjects and among pronoun types.

There are several issues that can be investigated concerning the factors that affect an ESL student's comprehension of anaphora, anaphoric relations and the determination of pronoun assignment. These include 1) the effects of proficiency level, 2) whether the task is in the recognition mode or the production mode, and whether 3) the number of possible antecedents, 4) the types of pronouns tested, and 5) the distance between anaphora and antecedent directly affect a subject's performance on pronoun assignment.

The purpose of this investigation was to answer five research questions, each of which can be translated into two statistically testable hypotheses: a null and an alternative hypothesis. The pairs of hypotheses will be listed in turn:

Question 1:

H_0: There will be no significant difference between ESL subjects from different proficiency levels on tests of anaphoric reference.

H_A: There will be a significant difference between ESL subjects from different proficiency levels on tests of anaphoric reference.

Question 2:

H_0: There will be no significant difference between ESL subjects' performance on recognition and production tests of anaphoric reference.

H_A: There will be a significant difference between ESL subjects' performance on recognition and production tests of anaphoric reference.

Question 3:

H_0: There will be no significant difference between ESL subjects' performance on a test of anaphoric reference with two antecedents and another test with four antecedents.

H_A: There will be a significant difference between ESL subjects' performance on a test of anaphoric reference with two antecedents and another test with four antecedents.

Question 4:

H_0: There will be no significant difference between ESL subjects' performance on subject, object, and possessive pronouns in tests of anaphoric reference.

H_A: There will be a significant difference between ESL subjects' performance on subject, object, and possessive pronouns in tests of anaphoric reference.

Question 5:

H_0: There will be no difference between ESL subjects' performance on items in which anaphora and antecedent are separated by 0-2 propositions and on items in which anaphora and antecedent are separated by 3-5 propositions.

H_A: There will be a significant difference between ESL subjects' performance on items in which anaphora and antecedent are separated by 0-2 propositions and on items in which anaphora and antecedent are separated by 3-5 propositions.

Method

Subjects. This research was conducted at the Center for English as a Second Language (CESL) at Southern Illinois University. The subject pool consisted of 50 full-time ESL students from three proficiency strata:

elementary (Level 1), n = 14; intermediate (Level 2), n = 25; advanced (Level 3), n = 11. The following native languages were represented: Arabic, Spanish, Malay, Japanese, Chinese, Farsi, Korean, and Twi.

Instrumentation. Miller's Tests of Anaphoric Reference (1973) were used as elicitation instruments. There were four tests: a 24-item multiple-choice recognition (MC 1) with two antecedents. Eleven items tested subject pronouns (pronoun type 1); 5 tested object pronouns (pronoun type 2); 4 tested possessive pronouns (pronoun type 7). Ten of the pronouns were separated from their antecedents by 0-2 propositions (prox A); the other ten were separated by 3-5 propositions (prox B). A second 24-item multiple-choice recognition test (MC 2) had four antecedents; it had 12 subject pronouns, 4 object pronouns, and 4 possessive pronouns. Of the total, 10 were prox A and 10 prox B. Each of these tests had 4 filler items.

There were two 50-item cloze tests, one with 10 non-pronoun items, the other with 11 such items; the purpose of these items, as in the multiple-choice tests, was to avoid "set." Cloze 1 tested 23 subject pronouns, 9 object pronouns, and 8 possessive pronouns. Twenty-one items were prox A (antecedent/anaphora separated by 0-2 propositions) and 19 prox B items (antecedent/anaphora separated by 3-5 propositions). Cloze 2 tested 22 subject pronouns, 9 object pronouns, and 8 possessive pronouns. Twenty were prox A and 19 prox B. Cloze 1 had 2 antecedents, Cloze 2 had 4.

These tests, as the author says, "contain four stories which were drawn from basal readers and modified so that the categories of pronouns were represented in proportion to their occurrence in the basal readers" (1973: Document Resume). Although the tests were aimed at the beginning English-speaking reader, we believe that the instrument is valid because

our foreign-student population demonstrate a parallel range from beginners in Level 1 to lower advanced in Level 3.

Analyses and Results

The summary statistics are presented in Table 1.

To answer Question 1 (the influence of proficiency level), we submitted the data to the Kruskal-Wallis test with chi-square approximation. The tabled value for the chi-square to reject the null hypothesis at the .05 level with 2 degrees of freedom is 5.99. The results were:

MC 1	$X^2 = 8.32$	$p < .05$
MC 2	$X^2 = 3.48$	n.s.
Cloze 1	$X^2 = 11.94$	$p < .05$
Cloze 2	$X^2 = 16.01$	$p < .05$

The finding is that there is a significant difference in the ranks of the three samples for MC 1, Cloze 1, and Cloze 2. There is no difference in the ranks of the three samples for MC 2. We retain the null hypothesis for MC 2, but we reject it and accept the alternative hypothesis for MC 1, Cloze 1 and Cloze 2. For these three tests, there is a difference between proficiency strata.

To answer Question 2 (recognition vs. production), we created two new variables by combining the MC tests into a recognition variable and the two cloze tests into a production variable. We subtracted the difference between the two and submitted the data to a t-test. The differences were significant for the three proficiency levels:

Level 1	$t = 12.84$	$p < .05$
Level 2	$t = 9.71$	$p < .05$
Level 3	$t = 5.13$	$p < .05$

There is a significant difference between our ESL subjects' performance on recognition and production items, the production items being much more difficult, and we reject the null hypothesis for Question 2.

To answer Question 3 (number of antecedents), we submitted the data from MC 1 with two antecedents and from MC 2 with four antecedents to a t-test. The resulting t statistic was t = .07 with 98 df. A similar procedure was conducted for Cloze 1 with two antecedents and Cloze 2 with four antecedents. The t statistic was t = .03 with 98 df. Both figures were statistically non-significant. For Question 3, the null hypothesis is therefore retained. The number of antecedents in the stories did not affect pronoun assignment.

Regarding Question 4, the data were submitted to the Kruskal-Wallis Test with chi-square approximation to determine whether there was a significant difference on the samples' performance on subject, object, and possessive pronouns. The results are:

MC 1	Type 1	$x^2 = 5.38$	n.s.
	2	$x^2 = 5.39$	n.s.
	7	$x^2 = 6.62$	$p < .05$
MC 2	Type 1	$x^2 = 3.83$	n.s.
	2	$x^2 = 1.12$	n.s.
	7	$x^2 = 1.74$	n.s.
Cloze 1	Type 1	$x^2 = 9.06$	$p < .05$
	2	$x^2 = 15.10$	$p < .05$
	7	$x^2 = 5.54$	n.s.
Cloze 2	Type 1	$x^2 = 6.33$	$p < .05$
	2	$x^2 = 18.55$	$p < .05$
	7	$x^2 = 18.03$	$p < .05$

The statistics are conclusive for Cloze 2, but not so for the other tests. We can conclude that for the cloze tests there is a significant difference in performance on the three different categories of pronouns but not for the MC tests. We therefore retain the null hypothesis for the MC tests but reject the null and accept the alternative hypothesis for the cloze tests.

To answer Question 5 (distance between pronoun and antecedent), we submitted the prox A items (antecedent/anaphora separated by 0-2 propositions) and the prox B items (antecedent/anaphora separated by 3-5 propositions) to a t-test for each test. The results are:

MC 1:	t = .10 df 18	n.s.
MC 2:	t = 0 df 18	n.s.
Cloze 1:	t = 2.36 df 38	$p < .05$
Cloze 2:	t = 1.74 df 37	$p < .05$

For the cloze tests the distance between antecedent and anaphora does make a difference, and we reject the null hypothesis. However, for the MC tests distance was not a factor influencing pronoun assignment, and we therefore retain the null hypothesis.

Being influenced by Bormuth et al. (1970) and Lesgold's (1974) research on the stability of difficulty issue, we conducted a post hoc analysis on the stability of difficulty ordering across proficiency levels. We submitted the data in Table 2 to the Coefficient of Concordance (W) with chi-square approximation. The purpose of the analysis is to obtain an overall index of the amount of agreement in the rank orders. We are not concerned here with the issue of the differences between the means for the three pronoun types but whether the rank orders of difficulty are consistent across the

proficiency strata. Except for the prox rank orders in MC 2, Cloze 1, and Cloze 2, there is no significance of agreement in the rank orders across the strata, i.e. none of the chi-square statistics were significant.

The same data are rearranged in Table 3 and Spearman rank-order correlations were calculated for the pairs of ranks for the recognition and production tasks. The results:

		Level 1	Level 2	Level 3
		rho	rho	rho
MC 1 & 2	1,2,7 ranks	1.*	-.5	-.13
Cloze 1 & 2	1,2,7 ranks	.67	1.*	1.*
MC 1 & 2	A,B	1.*	1.*	-1.*
Cloze 1 & 2	A,B	1.*	1.*	1.*

* significant

Discussion

The finding for Question 1 is not surprising because the more proficient subjects know more vocabulary, more grammar, and more reading strategies than the less proficient subjects. Consequently, the more proficient subjects are better equipped to utilize the graphemic-phonological, syntactic and semantic cues in the text and can reduce the redundancy faster. The two MC tests suffer from boundary effects as evidenced by the means in Table 1 and therefore do not exhibit discriminability or person separability as well as do the cloze tests.

The finding for Question 2 can be explained in a similar fashion. In addition to the boundary effects, the MC questions were nothing more than template-matching or simple word-recognition tasks. But the cloze procedure forced the subjects to read for both content and structure of

the texts, and therein lies its power. In sum, more cognitive and linguistic machinery were engaged by the cloze than by the MC tests.

The number of antecedents in each story did not have any perceived effect on pronoun assignment. Perhaps if a more complicated text had been utilized, or if ambiguity had been incorporated, the antecedent issue would have been different.

We believe that our data support Lesgold's (1974) finding: that there is variability in subjects' processing of anaphora and among different types of anaphoric pronouns. We think that if the results of the Kruskal-Wallis test for Question 5 are not convincing, Tables 2 and 3 definitely corroborate Lesgold's research.

The biggest surprise to us in this investigation was the finding that the distance between pronoun and antecedent did not matter in the MC tests. We have already pointed out that the MC tests have boundary effects for our samples and do not discriminate. However, the results from the cloze tests--that distance does have a determining effect--corroborate previous research. For example, de Beaugrande and Dressler (1983:60):

> Anaphora is the most common directionality for co-reference, since the identity of the conceptual content being kept current is made plain in advance. However, anaphora can still be troublesome if there is a lengthy stretch of text before the proform appears. By then, the original elements could have been displaced from active storage and other candidates may be mistakenly called.

Kintsch and Keenan (1973) have found that adding propositions to the base structures of sentences increases reading time. If those propositions

were added between pronoun and referent, we believe pronoun assignment would be affected. In addition, Garrod and Sanford (1977) reported that semantic distance between referent and antecedent affected processing time.

There are two pedagogical implications from this study: 1) some personal pronouns will be more difficult than others for some or all students, and, therefore 2) systematic instruction on personal pronouns will be necessary. This conclusion is supported by the scale of the means of the 3 levels, as shown below:

	Level 1 (n = 14)	Level 2 (n = 25)	Level 3 (n = 11)
MC 1	.90 (sd .19)	.93 (sd .17)	.97 (sd .09)
MC 2	.82 (.19)	.92 (.09)	.98 (.03)
Cloze 1	.51 (.23)	.65 (.15)	.78 (.14)
Cloze 2	.54 (.22)	.59 (.16)	.83 (.10)

Future research on ESL students' processing of anaphora might include such topics as the difference between "deep" and "surface" anaphora (cf. Hankamer and Sag, 1976); given and new information as it relates to anaphora (cf. Haviland and Clark, 1974); and the extent to which implicit causality in root verbs affects pronoun assignment (cf. Caramazza, Grober, and Garvey, 1977).

Table 1

Summary Statistics

Test	Number of Subjects	Number of Items	Mean of Raw Scores	SD	KR-20	50-item rtt
Multiple-Choice 1	50	24	22.18	3.86	.92	.96
Multiple-Choice 2	50	24	21.82	3.16	.84	.92
Cloze 1	50	50	32.00	9.49	.90	.90
Cloze 2	50	50	32.38	8.54	.89	.89

Table 2

Pronoun Type by Level

		Level 1		Level 2		Level 3	
		%	Rank	%	Rank	%	Rank
MC 1							
Prontype	1	91	1	92	3	97	3
	2	90	2	93	2	98	1.5
	7	89	3	94	1	98	1.5
Prox	A	91	1	93	1	96	2
	B	90	2	92	2	98	1
MC 2							
	1	85	1	94	2	100	1.5
	2	84	2	95	1	100	1.5
	7	71	3	85	3	98	3
	A	82	1	93	1	100	1
	B	81	2	92	2	99	2
Cloze 1							
	1	54	2	67	1	79	2
	2	34	3	60	3	74	3
	7	58	1	64	2	80	1
	A	56	1	69	1	81	1
	B	44	2	59	2	74	2
Cloze 2							
	1	61	1	75	1	81	2
	2	39	3	59	3	80	3
	7	51	2	71	2	89	1
	A	60	1	75	1	86	1
	B	49	2	65	2	78	2

Table 3

Pronoun Type by Test

		MC 1		MC 2		Cloze 1		Cloze 2	
		Level 1							
		%	Rank	%	Rank	%	Rank	%	Rank
Prontype	1	91	1	85	1	54	2	61	1
	2	90	2	84	2	34	3	39	3
	7	89	3	71	3	58	1	51	2
Prox	A	91	1	82	1	56	1	60	1
	B	90	2	81	2	44	2	49	2
		Level 2							
	1	92	3	94	2	67	1	75	1
	2	93	2	95	1	60	3	59	3
	7	94	1	85	3	64	2	71	2
	A	93	1	93	1	69	1	75	1
	B	92	2	92	2	59	2	65	2
		Level 3							
	1	97	3	100	1.5	79	2	81	2
	2	98	1.5	100	1.5	74	3	80	3
	7	98	1.5	98	3	80	1	89	1
	A	96	2	100	1	81	1	86	1
	B	98	1	99	2	74	2	78	2

REFERENCES

de Beaugrande, Robert, and Wolfgang Dressler. 1983. Introduction to text linguistics. London: Longman Group Limited.

Bormuth, John R., Julian Carr, John Manning, and David Pearson. 1970. Children's comprehension of between- and within-sentence syntactic structures. Journal of Educational Psychology 61(5):349-357.

Caramazza, Alfonso, Ellen Graber, and Catherine Garvey. 1977. Comprehension of anaphoric pronouns. Journal of Verbal Learning and Verbal Behavior 16(5):501-609.

Fishman, Anne Stevens. 1978. The effects of anaphoric references and noun phrase organizers on paragraph comprehension. Journal of Reading Behavior 10(2):159-170.

Garrod, Simon, and Anthony Sanford. 1977. Interpreting anaphoric relations: the integration of semantic information while reading. Journal of Verbal Learning and Verbal Behavior 16(1):77-90.

Garvey, Catherine, Alfonso Caramazza, and Jack Yates. 1976. Factors influencing assignment of pronoun antecedents. Cognition 3(3):227-243.

Hankamer, Jorge, and Ivan Sag. 1976. Deep and surface anaphora. Linguistic Inquiry 7(3):391-426.

Haviland, Susan E., and Herbert H. Clark. 1974. What's new: acquiring new information as a process in comprehension. Journal of Verbal Learning and Verbal Behavior 13(4):512-521.

Kintsch, Walter, and Janice Keenan. 1973. Reading rate and retention as a function of the number of propositions in the base structure of sentences. Cognitive Psychology 5(3):257-274.

Lesgold, Alan M. 1974. Variability in children's comprehension of syntactic structures. Journal of Educational Psychology 66(3):333-338.

Miller, Larry A. 1973. Tests of anaphoric reference - multiple choice format (TAR-MC) and tests of anaphoric reference - cloze format (TAR-C). ERIC Reports ED 091 768. Washington, D.C.: Educational Resources Information Center.

Richek, Margaret Ann. 1976-77. Reading comprehension of anaphoric forms in varying linguistic contexts. Reading Research Quarterly 12(2):145-165.

The Authors

Charles Parish is Professor of Linguistics/ESL at Southern Illinois University, Carbondale, and former Academic Director of CESL. His current interests are in innovative methods in EFL/ESL and in notional/functional syllabuses. He has been a Fulbright lecturer in Burma and in Italy.

Kyle Perkins is Associate Professor of Linguistics/ESL at Southern Illinois University, Carbondale. He is co-editor, with John Oller, of two books on testing, and he has recently returned from a Fulbright year in Egypt.